

MODELOS ARIMA PARA LA TEMPERATURA MEDIA ANUAL EN SEIS CIUDADES ESPAÑOLAS

S. Pérez Boada *, J. Gallego-Díaz Fajardo **, F. Elías Castillo **

* Sección de Estadística. Área de Informática Científica INIA

** Servicio de Investigación. Unidad Investigadora de Ecología, Comunidad de Madrid

Resumen

Se ha realizado el análisis univariable de las series de temperatura media anual de seis ciudades españolas, utilizando los modelos ARIMA de Box-Jenkins. La serie de Badajoz fue un AR(2) y la de Barcelona un AR(1). Las cuatro restantes, Madrid, La Coruña, San Fernando (Cádiz) y Valencia, resultaron ser un ARIMA (0, 1, 1). Para todas ellas, se ha obtenido la predicción para el año 1989.

Palabras clave

Análisis Univariable de Series Temporales, Método Box-Jenkins, Modelos Arima, Temperatura Media Anual.

Introducción

La característica más importante de los datos de series temporales es la dependencia entre las observaciones correspondientes a sucesivos períodos de tiempo igualmente espaciados. El análisis de esta situación se puede realizar en el dominio de las frecuencias, dando lugar al Análisis Espectral, o en el dominio del tiempo, según el enfoque introducido por Box y Jenkins (1976).

En el primero de ellos, las observaciones se expresan como suma de ondas sinusoidales a diferentes frecuencias o períodos, mientras que el análisis de Box-Jenkins utiliza los modelos ARIMA (siglas inglesas de Autorregresivos Integrados de Medias Móviles), que expresan de mane-

ra sencilla la relación entre la variable estudiada en el período actual y los períodos anteriores.

A lo largo de la historia, se han aplicado las series temporales en una gran variedad de ciencias, como Economía, Sociología, Hidrología, Ecología, etc., siendo en la primera de ellas donde más se han utilizado. En el caso de la Climatología, y más concretamente en el estudio de las temperaturas, existen diversos trabajos de aplicación, que utilizan, en su mayoría, el Análisis Espectral, entre los que se citan Pruscha (1986), Madden (1977), Hancock y Yarger (1979), Olsen y Mikkelsen (1983) y Persaud y Chang (1985).

En este trabajo se estudia la temperatura media anual en seis ciudades: Badajoz, Barcelona, La Coruña, Madrid, San Fernando (Cádiz) y Valencia. El objetivo es modelizar cada una de estas series y predecir a corto plazo. El método que se va a seguir es el análisis univariable según el enfoque de Box y Jenkins.

Material y Métodos

Las observaciones representan la temperatura media anual obtenida de la siguiente forma: primero se considera la semisuma de las temperaturas máxima y mínima de cada día; a continuación, se calcula la media mensual de dichas temperaturas diarias; por último, se obtiene la media de los doce meses. Las temperaturas vienen dadas en grados centígrados.

La fuente de los datos han sido los listados del Instituto Nacional de Meteorología. las figuras 1 a 6 representan las series de datos originales para cada ciudad, junto con el número de observaciones disponible (N), la media (μ), la desviación típica (s), el coeficiente de variación (CV) y los años considerados.

Todas las series constan de más de cien años, excepto el caso de Valencia en que sólo se dispone de 87 observaciones. Es en esta última ciudad donde se observa el único dato atípico (fuera del intervalo: media + tres desviaciones típicas) en las series originales, en el año 1919, con un valor de 14,9°C. Todas las series están completas, excepto Madrid, donde faltan las observaciones correspondientes a los meses de febrero de 1875, noviembre y diciembre de 1936, enero, febrero y abril de 1937 y marzo y abril de 1939. Estos datos se rellenaron con la media del mes correspondiente calculada a partir de los años disponibles. A continuación, se obtuvo el dato anual como media de los doce meses.

Todos los cálculos se han realizado en el ordenador de la Sección de Proceso de Datos del INIA, un CDC-CYBER 180, utilizando el programa BMDP2T.

El método que se ha utilizado para el análisis univariable de las series de datos elegidas ha sido el de Box y Jenkins. El objetivo principal de este método es obtener un modelo paramétrico sencillo que explique la estructura de dependencia lineal de las observaciones y sirva para predecir valores futuros. Este método es adecuado para series de tiempo estacionarias. Se dice que una serie Y_t , $t=1,2,\dots,N$ es estacionaria cuando su media y varianza son constantes en el tiempo y la covarianza entre las observaciones Y_t e Y_{t-k} , $k=1,2,\dots,N-1$, depende sólo del valor del retardo k .

Los modelos introducidos por Box y Jenkins pueden ser de dos tipos: modelos autorregresivos (AR) y modelos de media móvil (MA). En los modelos autorregresivos de orden p , $AR(p)$, la observación en el tiempo t , Y_t , se escribe como

una combinación lineal de p observaciones anteriores:

$$Y_t - \mu = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

o bien

$$\phi_p(B) (Y_t - \mu) = e_t,$$

donde:

— $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, siendo B un operador lineal de retardo tal que $B^r Y_t = Y_{t-r}$.

— e_t es una serie de variables aleatorias independientes y con distribución Normal de media cero y varianza σ^2 constante, comúnmente llamada "ruido blanco".

— μ es la media de la serie $\{Y_t\}$.

Por su parte, en los modelos de media móvil de orden q , $MA(q)$, la observación en el tiempo t se expresa en función de la serie $\{e_t\}$ de la siguiente forma:

$$Y_t - \mu = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t$$

o bien

$$Y_t - \mu = \theta_q(B) e_t,$$

con

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

Los modelos autorregresivos y de media móvil se pueden combinar dando lugar a un modelo mixto, $ARMA(p,q)$, con la siguiente expresión:

$$\phi_p(B) (Y_t - \mu) = \theta_q(B) e_t$$

Por otra parte, una condición necesaria y suficiente para que la serie sea estacionaria es que las raíces del polinomio $\phi(B)$ sean todas mayores que uno en valor absoluto. Cuando la serie no es estacionaria y $\phi(B)$ tiene d raíces unitarias, se utilizan los modelos autorregresivos integrados de media móvil, $ARIMA(p,d,q)$, cuya expresión es:

$$\phi_p(B) \nabla^d Y_t = \theta_q(B) e_t$$

siendo $\nabla = 1 - B$ el operador de diferencias (Peña, 1987a, vol. II, pág. 535).

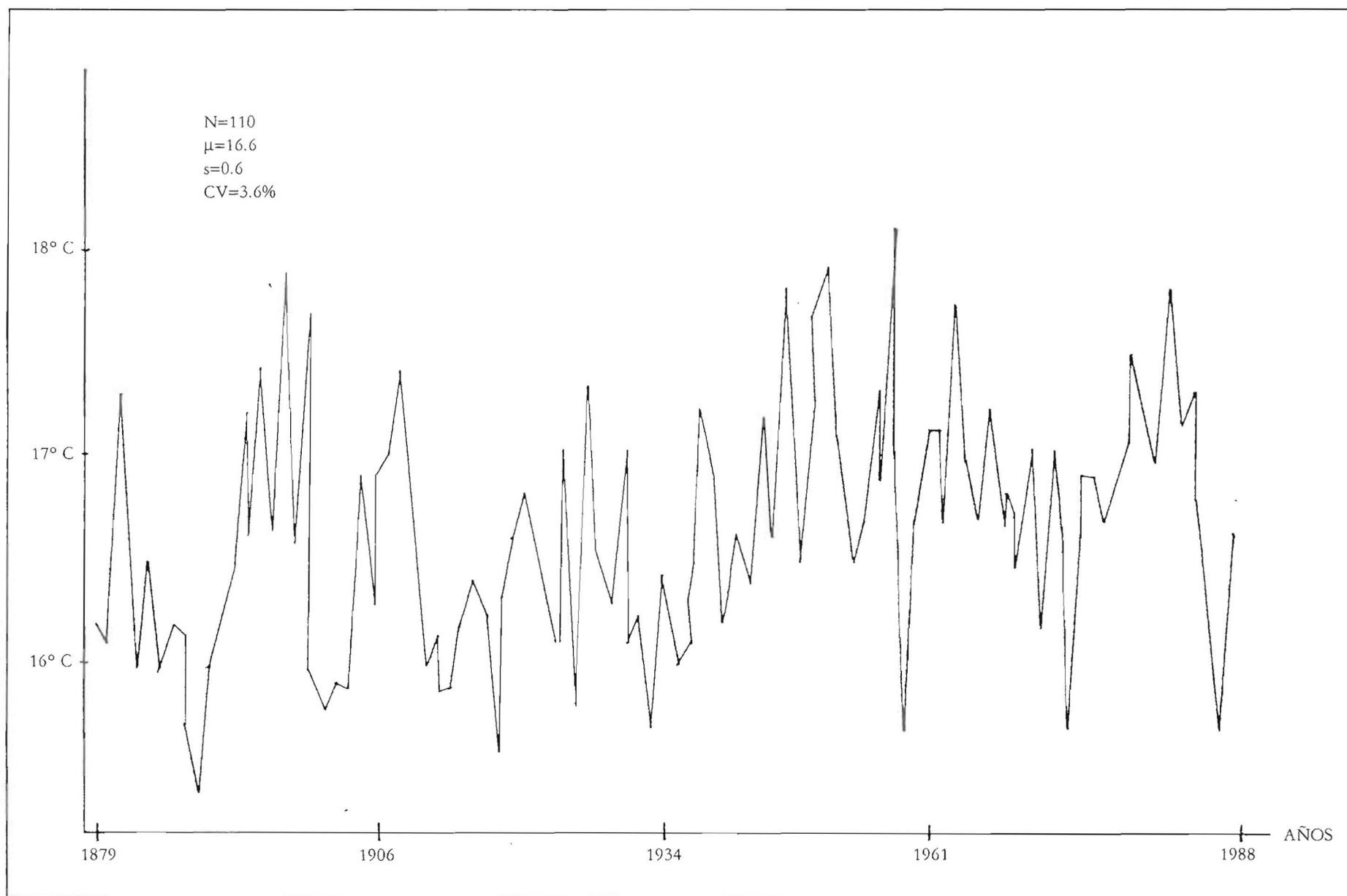


Figura 1.—Temperatura media anual en Badajoz. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

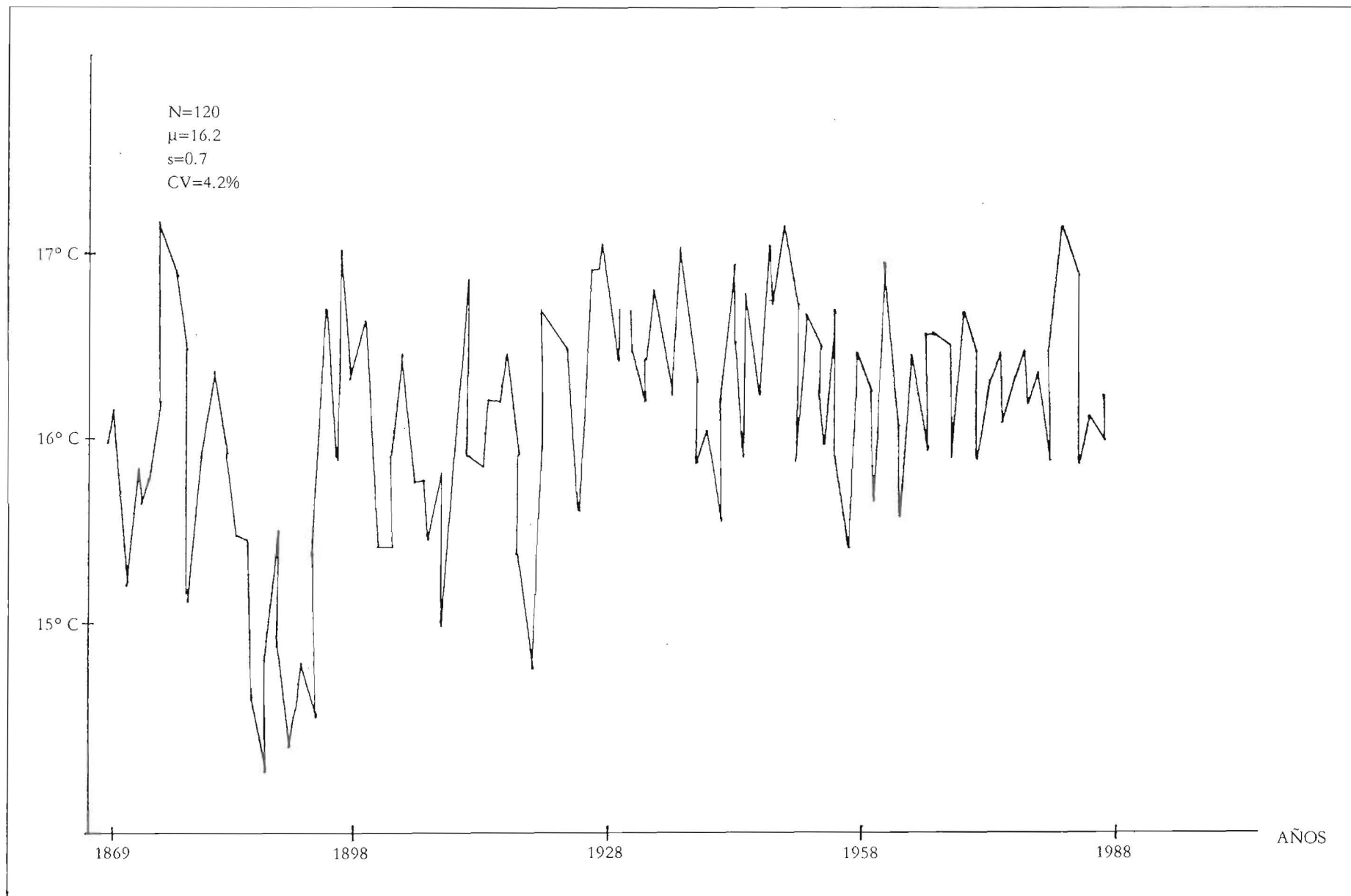


Figura 2.—Temperatura media anual en Barcelona. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

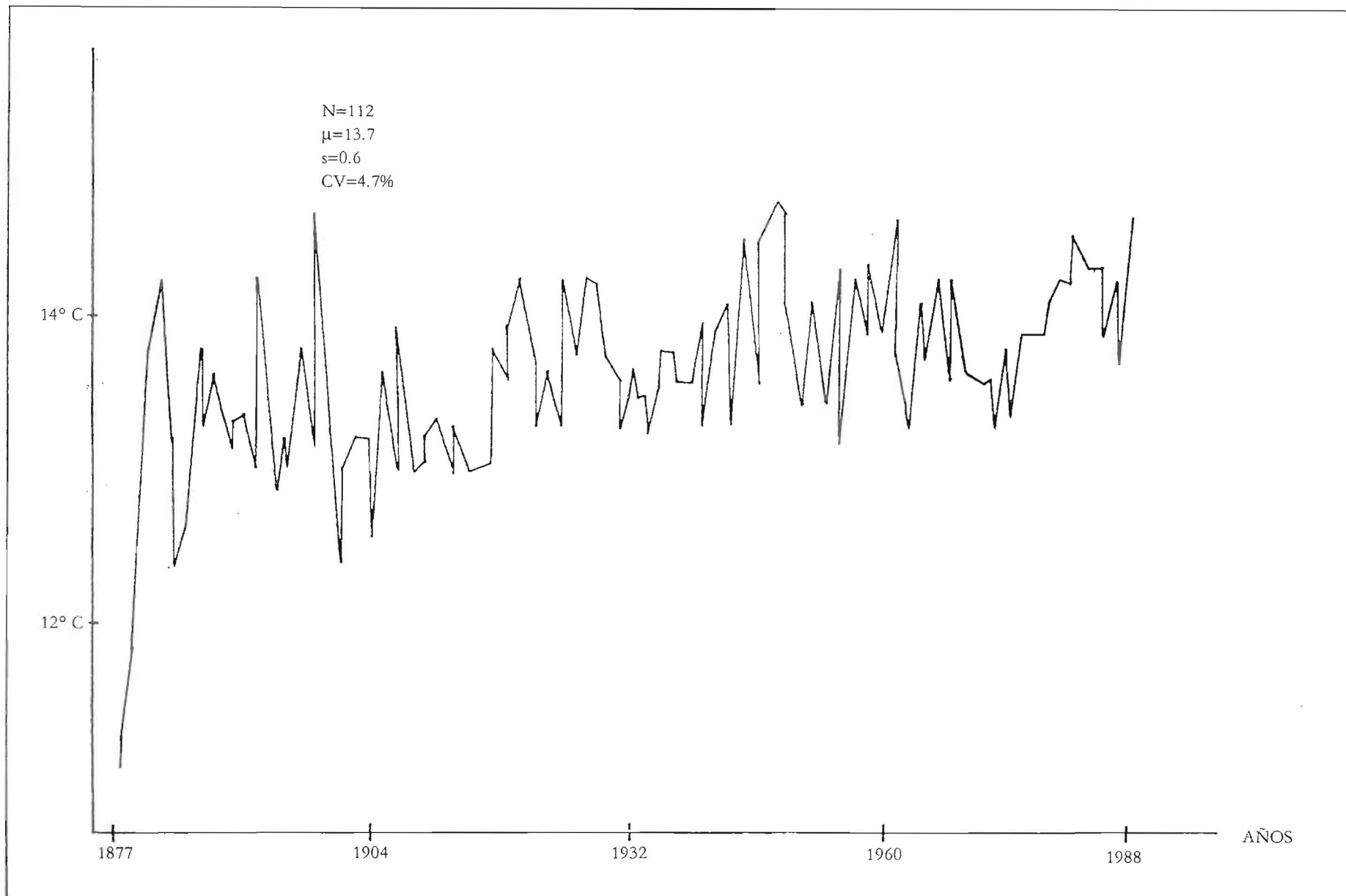


Figura 3.—Temperatura media anual en La Coruña. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

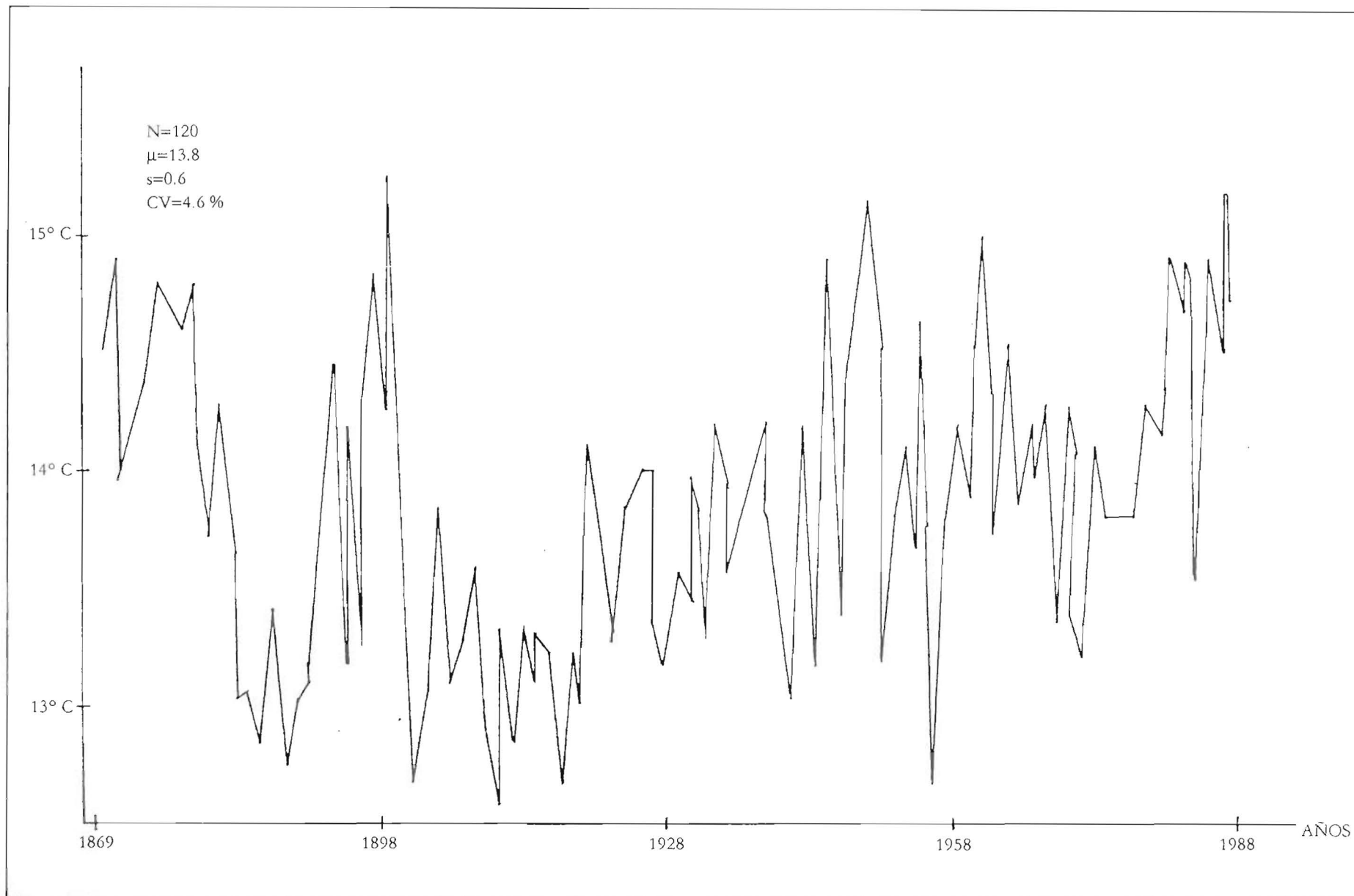


Figura 4.—Temperatura media anual en Madrid. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

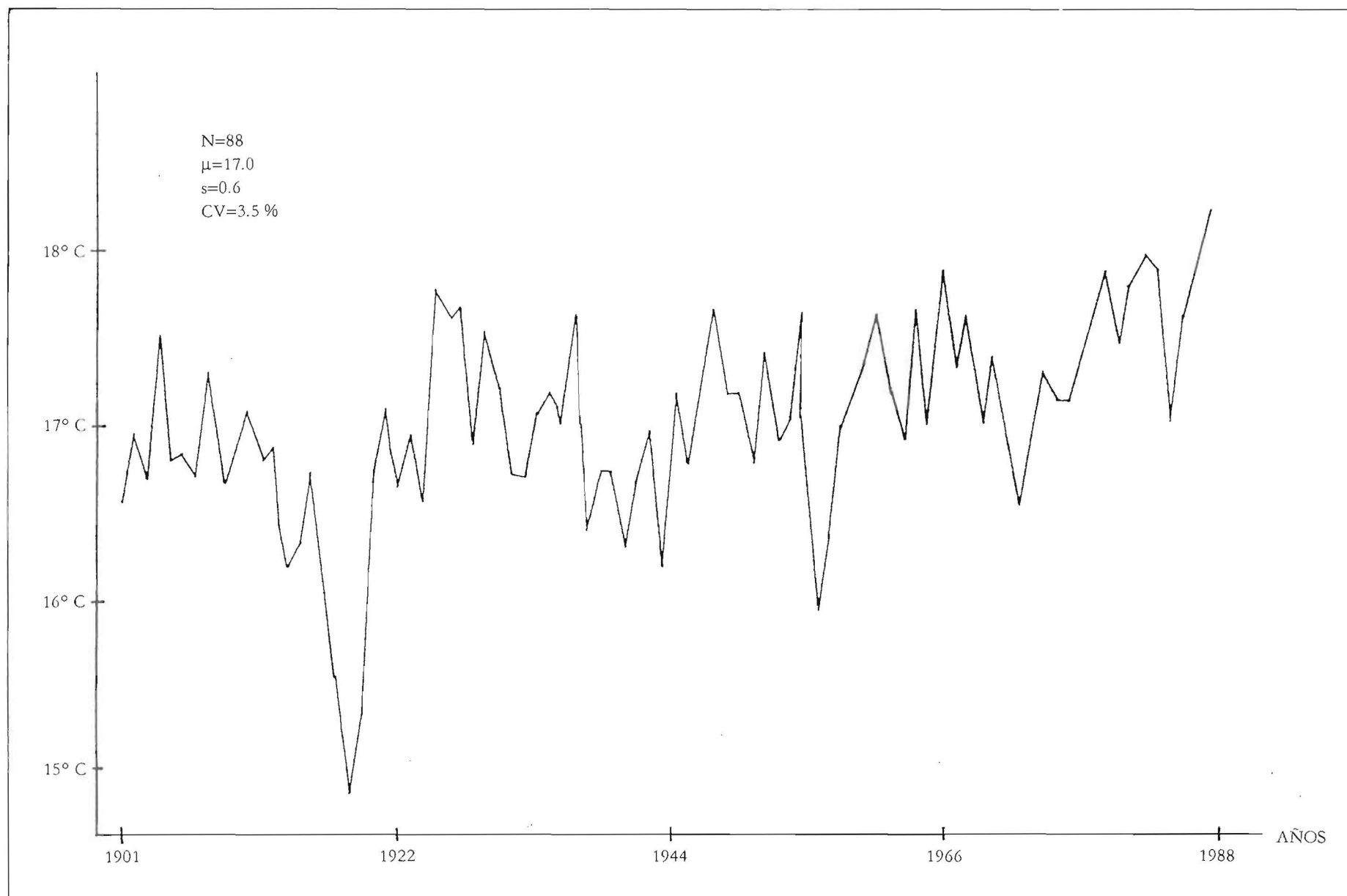


Figura 6.—Temperatura media anual en Valencia. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

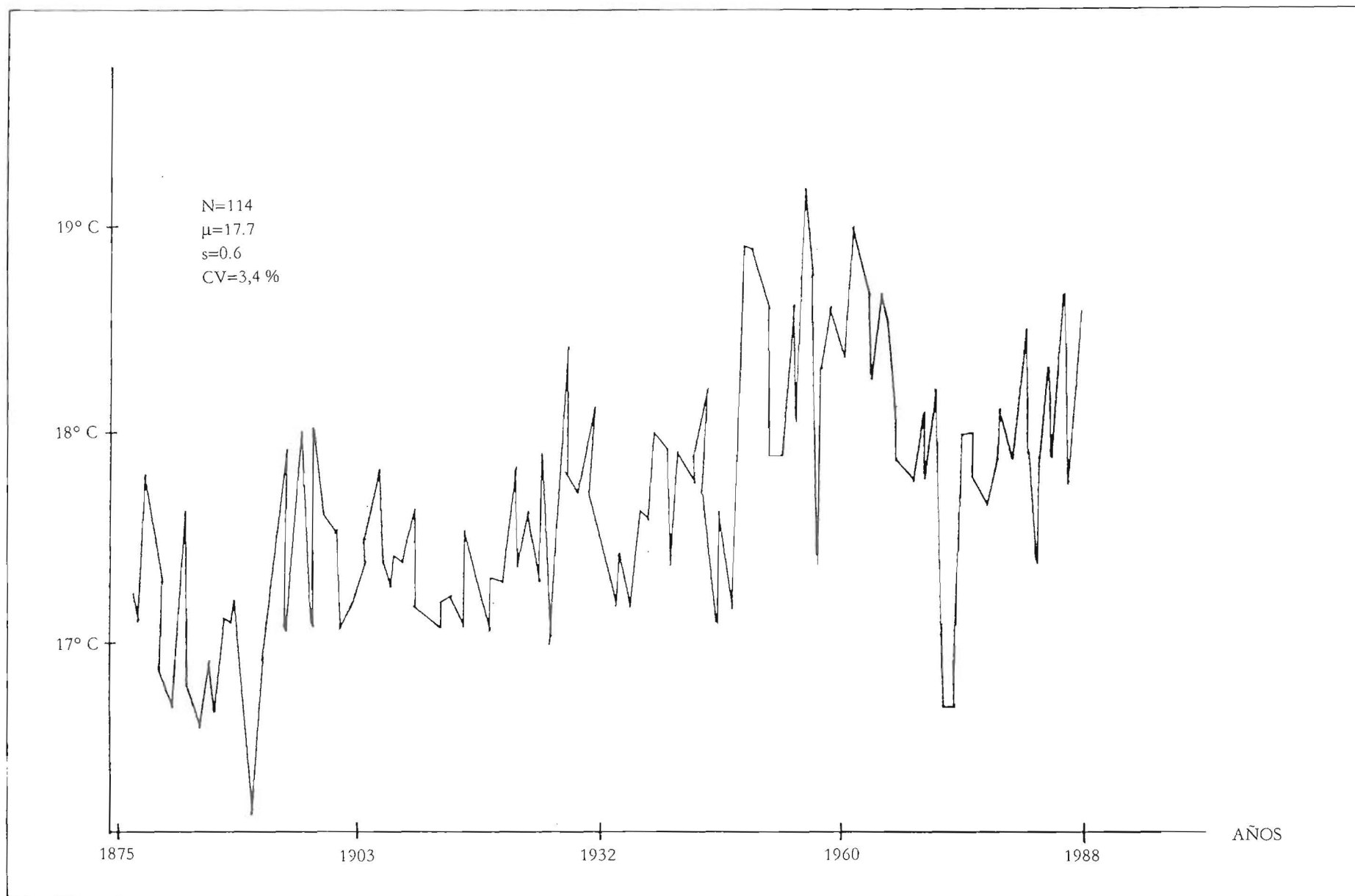
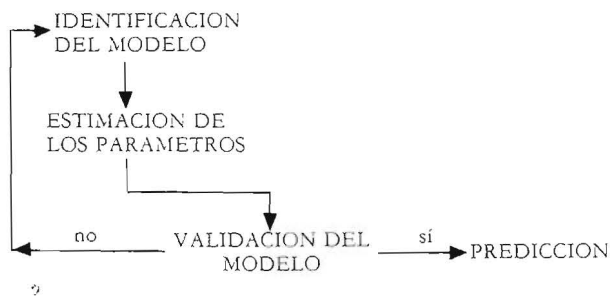


Figura 5.—Temperatura media anual en San Fernando. N = número de años considerado, μ = media, s = desviación típica, CV = coeficiente de variación.

La ventaja de los modelos ARIMA es que con un pequeño número de parámetros estimados a partir de los datos, se pueden describir una gran variedad de series temporales.

El método Box-Jenkins es un procedimiento iterativo de construcción del modelo en base a los datos, que consta de tres etapas resumidas en el siguiente esquema:



— En la etapa de identificación del modelo se han tenido en cuenta el gráfico de la serie original y las funciones de autocorrelación simple y parcial de la muestra (f.a.s. y f.a.p.). En dicho gráfico se pueden detectar datos atípicos, tendencias y/o falta de homogeneidad de varianza en

las observaciones. Si se detecta algún valor atípico en los datos originales, es conveniente restar su efecto para que no falsee la identificación del modelo. Si existe una tendencia definida en los datos, por lo que la serie no es estacionaria, hay que, o bien ajustar dicha tendencia mediante un modelo de regresión, o bien diferenciar la serie. Si se observa falta de homogeneidad de varianza, se pueden utilizar las transformaciones de Box-Cox (1964).

Por otra parte, las funciones de autocorrelación parcial y simple son la representación de los coeficientes de autocorrelación en función del retardo k y miden la relación lineal entre observaciones separadas k períodos, con independencia o no de los valores intermedios respectivamente. La forma de estas funciones sirve para identificar el tipo y el orden del modelo. Así por ejemplo, en la figura 7 están representadas la f.a.s. y f.a.p. de la serie de Badajoz. Como la f.a.s. decrece exponencialmente a la f.a.p. tiene valores significativos en los retardos 1 y 2, se puede postular un modelo AR(2).

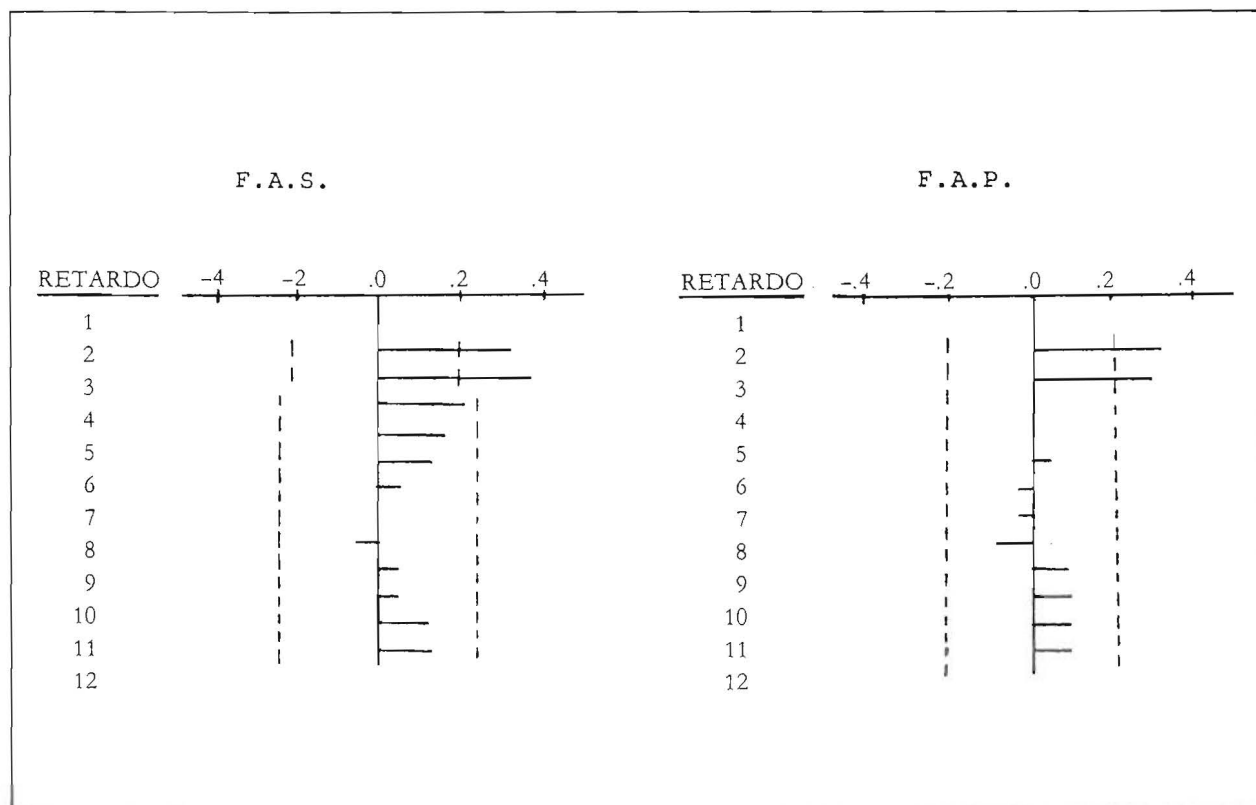


Figura 7.—Función de autocorrelación simple (F.A.S.) y función de autocorrelación parcial (F.A.P.) para la serie de Badajoz. Las líneas de puntos representan los límites del intervalo de confianza.

— En la segunda etapa, los parámetros del modelo postulado se han estimado según el método “backcasting” o mínimos cuadrados incondicionados (Box y Jenkins, 1976, pág. 213).

— En la etapa de validación se estudiaron:

1. La serie de residuos para ver si existían valores atípicos y para comprobar que era ruido blanco. Se consideraron datos atípicos aquellos que estaban fuera del intervalo $\pm 3\sigma_e$, siendo σ_e la estima de la desviación típica de los residuos, (Chang et al., 1988). Por otra parte, para verificar que eran ruido blanco se utilizó una prueba de Student para ver que tenían media nula, y el análisis del gráfico de la serie frente al tiempo para comprobar la estabilidad de la varianza. La incorrelación de los residuos se comprobó viendo si alguno de los valores de su f.a.s. y su f.a.p. era significativamente distinto de cero al 95 % y mediante el estadístico de Ljung-Box, Q_n , que sirve para probar la incorrelación de los n primeros residuos (Ljung y Box, 1978).

Por otra parte, se realizó una prueba de normalidad de la serie, calculando el estadístico chi-cuadrado y los valores de los coeficientes de asimetría y kurtosis.

2. Sobreajustes del modelo para ver que no existía otro del mismo tipo de orden superior que fuese mejor que el postulado (Peña, 1987a, vol.II, pág. 561).

3. La estabilidad del modelo a lo largo del tiempo. Para ello, se estimaron los parámetros del mismo, considerando la primera y segunda mitad de la serie por separado, y se comprobó que no existían diferencias significativas.

— Una vez validado el modelo, se obtuvieron las predicciones para el año 1989.

El tratamiento que se dio a los datos atípicos fue el siguiente: una vez detectados en la fase de validación, se incluyeron en el modelo aplicando el análisis de intervención (Box y Tiao, 1975), con el fin de contrarrestar su efecto. Se trata de un modelo muy simple donde la variable de in-

tervención, X_t , es una variable binaria que vale cero en todos los años, excepto en el año en cuestión, donde vale uno. De esta forma, resulta un modelo del tipo:

$$\phi_p(B) \nabla^d (Y_t - \beta X_t) = \theta_q(B) e_t$$

donde la estima del parámetro β está relacionada con la diferencia entre el valor del dato atípico y la media de la serie (Peña, 1987b).

Resultados y discusión

En la tabla 1 están los resultados obtenidos para todas las series. En la primera columna se da la ciudad y el modelo ARIMA correspondiente. La segunda columna presenta el desarrollo de la ecuación del modelo, con las estimas de los parámetros y su correspondiente desviación típica. Todas las series están centradas. En el caso de Badajoz, se da también el coeficiente de correlación entre los parámetros estimados. Los modelos de La Coruña, Madrid y San Fernando, contienen un parámetro adicional “c” debido a las intervenciones en los datos atípicos existentes en los residuos del ajuste.

En las siguientes cinco columnas están los estadísticos de validación del modelo. En primer lugar, el estadístico de Ljung-Box, Q_n , que hay que comparar con una chi-cuadrado con $n-p$ grados de libertad, siendo p el número de parámetros estimados. A continuación, están la media y los coeficientes de asimetría y kurtosis (este último restado de 3) de la serie de residuos, junto con sus desviaciones típicas. En la siguiente columna se da la raíz cuadrada del cuadrado medio residual del ajuste. Por último se dan las predicciones para el año 1989 junto con sus desviaciones típicas. Dichos errores son bajos dado que los datos de temperaturas medias anuales tienen una variabilidad pequeña, como lo indica su coeficiente de variación que en ningún caso supera el 5 %.

En el caso de Badajoz se consiguió un AR(2), mientras que para Barcelona, el modelo obtenido fue un AR(1). Para las demás ciudades, La

TABLA I
 MODELOS PARA LAS SEIS SERIES DE TEMPERATURA MEDIA ANUAL
 ESTADISTICOS DE VALIDACION Y PREDICCIONES PARA EL AÑO 1989

Ciudad y modelo	Ecuación del modelo	Validación					s	Predicción año 1989
		Q ₁₂	Q ₂₄	\bar{e}	A	K		
Badajoz ARIMA (2, 0, 0)	$Y_t = 7,88 + 0,23Y_{t-1} + 0,31Y_{t-2} + e_t$, corr= 0,33 (0,09) (0,09)	3,8	9,4	-0,0013 (0,0509)	-0,09 (0,23)	-0,58 (0,47)	0,54	16,5 (0,5)
Barcelona ARIMA (1, 0, 0)	$Y_t = 7,76 + 0,53Y_{t-1} + e_t$ (0,08)	12	28	0,0012 (0,0529)	-0,20 (0,22)	-0,32 (0,45)	0,58	16,3 (0,6)
La Coruña ARIMA (0, 1, 1)	$Y_t = c + Y_{t-1} - 0,84e_{t-1} + e_t$, $\begin{cases} -2,01, t=1877 \\ c= 2,01, t=1880 \\ 0, \text{ resto} \end{cases}$ (0,05)	11	27	0,046 (0,042)	0,07 (0,23)	0,13 (0,47)	0,46	14,2 (0,5)
Madrid ARIMA (0, 1, 1)	$Y_t = c + Y_{t-1} - 0,65e_{t-1} + e_t$, $\begin{cases} -1,07, t=1901 \\ c= 1,07, t=1902 \\ 0, \text{ resto} \end{cases}$ (0,05)	15	21	0,005 (0,048)	-0,09 (0,22)	0,01 (0,45)	0,54	14,7 (0,5)
San Fernando ARIMA (0, 1, 1)	$Y_t = c + Y_{t-1} - 0,78e_{t-1} + e_t$, $\begin{cases} -1,34, t=1971 \\ c= 1,34, t=1973 \\ 0, \text{ resto} \end{cases}$ (0,06)	12	24	0,040 (0,039)	0,35 (0,23)	0,19 (0,46)	0,42	18,2 (0,4)
Valencia ARIMA (0, 1, 1)	$Y_t = c + Y_{t-1} - 0,52e_{t-1} + e_t$ (0,09)	10	22	0,031 (0,051)	-0,17 (0,26)	0,45 (0,52)	0,48	18,1 (0,5)

NOTAS: Y_t es la temperatura; e_t es ruido blanco; Q_n es el estadístico de Ljung-Box; \bar{e} , A y K son respectivamente la media, asimetría y kurtosis de los residuos; s es la raíz del cuadrado medio del error; los números entre paréntesis son los errores estándar de las estimas.

Coruña, Madrid, San Fernando y Valencia, el modelo resultante fue el mismo: un MA(1) con una diferencia. Se trata de modelos muy sencillos, casi todos ellos con un solo parámetro, debido a que los datos utilizados son medidas anuales, lo que enmascara el efecto estacional.

Sería pues interesante analizar también las series mensuales de temperatura, punto que será objeto de un próximo trabajo para el caso de Madrid.

En ninguna de las series estudiadas se han detectado ciclos de varios años, que impliquen un comportamiento periódico de las temperaturas. Lo que sí se ha observado es una tendencia creciente en las mismas, sobre todo a partir de principios de este siglo, como se aprecia en la figura 8 en la que se han representado las medias móviles de amplitud 10, centradas cada dos años, para las seis ciudades. Aunque las temperaturas siguen la pauta de variación conocida para el clima peninsular durante los últimos 120 años, con ciertos intervalos fríos o cálidos, como por ejem-

plo entre los años 1876 y 1883, la gráfica evidencia además una clara, aunque pequeña, elevación de las temperaturas. Las causas de este paulatino aumento son discutibles, pero una de las más importantes puede ser el efecto de la ciudad y el hombre sobre el clima. Las condiciones climáticas en las ciudades difieren de las de su entorno, tanto más cuanto mayor sea el núcleo urbano. Entre los factores que más afectan al clima de ciudad pueden destacarse los siguientes:

- Emisión de calor debida a distintos tipos de actividad económica y cuya cuantía, en algunos casos, es comparable a la aportada por la radiación solar incidente sobre el área de la ciudad.
- El albedo de la superficie, generalmente menor en las zonas urbanas que en las rurales.
- La evaporación, que suele ser sensiblemente menor dentro del recinto urbano.

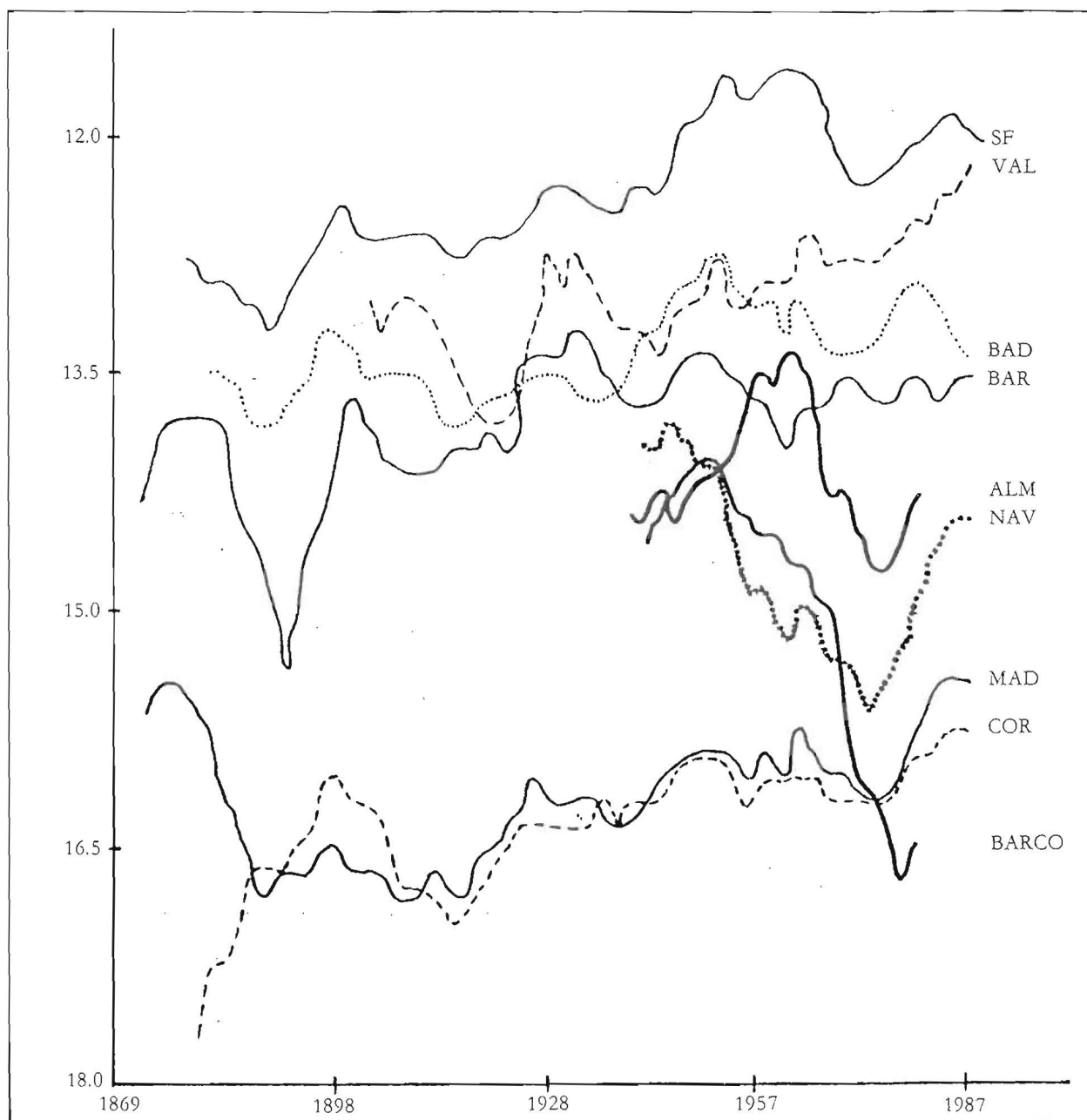


Figura 8.—Medias móviles de las temperaturas medias anuales de ocho ciudades españolas: S. Fernando (SF), Valencia (VAL), Badajoz (BAD), Barcelona (BAR), Almadén (ALM), Navacerrada (NAV), Madrid (MAD), La Coruña (COR) y Barco de Avila (BARCO).

Nota: Las temperaturas de Navacerrada y Barco de Avila corresponden a las originales habiéndose sumado 9° C y 4° C respectivamente.

También debe mencionarse la mayor rugosidad de la superficie dentro de la ciudad, y la contaminación atmosférica debida a emisiones sólidas, líquidas o gaseosas producidas por el hombre. La importancia de todos estos factores varía con la estación y el tipo de ciudad.

Uno de los rasgos principales del clima de ciudad es la "isla de calor", resultado de los tres

efectos mencionados anteriormente, que se caracteriza por temperaturas del aire más elevadas que en las áreas rurales.

Los resultados de varios investigadores han demostrado que una "isla de calor" suele tener una estructura compleja. La temperatura del aire en una gran ciudad es generalmente uno o dos grados centígrados superior a las de las zonas cir-

cundantes, si bien durante la noche y bajo condiciones de calma y con vientos ligeros, esta diferencia puede alcanzar hasta seis u ocho grados.

Todo parece indicar que el efecto de la “isla de calor” es más acusado durante la noche y, que el aumento de temperatura en las ciudades con respecto al régimen térmico de su entorno depende principalmente del aporte de calor de origen antropogénico y que las variaciones en el contenido de aerosoles afectan poco a las temperaturas.

A modo de comparación, se han incluido también en la figura 8 las gráficas de las medias móviles correspondientes a tres localidades no urbanas: Almadén, Navacerrada y Barco de Avila, en el intervalo común (1941-1980). Se observa que en el período considerado las oscilaciones de la temperatura media anual son mucho más acusadas en estas tres localidades que en el resto de las ciudades, donde el efecto del impacto del hombre ejerce una mayor influencia en el régimen térmico.

Comparando en dichas gráficas los años comunes de observación de las tres localidades, en las que se supone pequeño el efecto de la “isla de calor”, con las seis estaciones restantes, puede verse una oscilación extrema de temperatura superior a 1,5°C en las primeras, mientras que en las segundas es inferior a 1°C.

Agradecimientos

Este trabajo está incluido dentro del Proyecto n.º 7508 con título “Estudio estadístico de las series pluviométricas y termométricas más largas de la Península Ibérica”, de la Comunidad de Madrid, financiado por el INIA.

Summary

ARIMA models for the annual mean temperature in six spanish cities

Univariate analysis of annual mean temperature series for six spanish cities has been perfor-

med. Box-Jenkins ARIMA models have been used. Badajoz time series was an AR(2) while Barcelona was an AR(1). The four remaining ones, Madrid, La Coruña, San Fernando (Cádiz) and Valencia, turned out to be an ARIMA (0, 1, 1). Forecasts for the year 1989 were obtained for all series.

Key words: Univariate Time Series Analysis, Box-Jenkins Method, ARIMA Models, Annual Mean Temperature.

Referencias bibliográficas

- BOX, G. E. P., COX, D. R. (1964): “An analysis of transformations”. *Jour. Royal Statistical Soc.*, B26, 211-252.
- BOX, G. E. P., JENKINS, G. M. (1976): “Time series analysis”. Forecasting and control. Revised Edition. Holden-Day, California. 575 pp.
- BOX, G. E. P., TIAO, G. C. (1975): “Intervention analysis with applications to economic and environmental problems”. *Jour Amer. Stat. Assoc.*, 70 (march), 70-79.
- BUDYKO, M. I. (1982): “The earth’s climate: past and future”. International Geophysics Series, vol. 29. Academic Press, 307 pp.
- CHANG, I., TIAO, G. C., CHEN, C. (1988): “Estimation of time series parameters in the presence of outliers”. *Technometrics*, 30, 2, 193-204.
- GEIGER, R. (1966): “The climate near the ground”. Harvard University Press. Cambridge. Mass.
- HANCOCK, D. J., YARGER, D. N. (1979): “A cross-spectral analysis of sunspots and monthly mean temperatures and precipitation for the contiguous United States”. *Journal of the Atmospheric Sciences*, 36, 746-753.
- HARRINGTON, J. B. (1987): “Climate change: a review of causes”. *Canadian Journal of Forest Research*, 1313-1339.
- LJUNG, G. M., BOX, G. E. P. (1978): “On a measure of lack of fit in time series models”. *Biometrika*, 65, 297-303.
- MADDEN, R. A. (1977): “Estimates of the autocorrelations and spectra seasonal mean temperatures over North America”. *Monthly Weather Review*, 105, 9-18.
- OLESEN, J. E., MIKKELSEN, S. A. (1983): “Description and prediction of night minimum temperature by means of statistical models”. *Tidsskrift for plantavl*, 87, 2, 97-110.

- PEÑA, D. (1986): "Estadística. Modelos y Métodos, 1. *Fundamentos*. Alianza Ed. Madrid. 402 pp.
- PEÑA, D. (1987a): "Estadística. Modelos y Métodos, 2. *Modelos lineales y series temporales*. Alianza Ed. Madrid. 691 pp.
- PEÑA, D. (1987b): "Measuring the importance of outliers in ARIMA models. *New Perspectives in Theoretical and Applied Statistics*. Edited by Dr. M. L. Puri et al. John Wiley and Sons, Inc.
- PERSAUD, N., CHANG, A. C. (1985): "Time series analysis of daily solar radiation and air temperature measurements for use in computing potential evapotranspiration". *Transactions of the A.S.A.E. (USA)* 28, 2, 462-470.
- PRUSCHA, H. (1986): "A note on time series analysis of yearly temperature data". *Jour. Royal Statistical Soc., A*, 149, 174-185.